Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors

Sangjoon Lee[†], Jinyoung Kim[†], and Gerald Marschke[††]
[†]SUNY at Buffalo, [††]SUNY at Albany

January 11, 2004

Abstract

It has long been recognized that worker wages and possibly productivity are higher in large firms. Moreover, at least since Schumpeter (1942) economists have been interested in the relative efficiency of large firms in the research and development enterprise. This paper uses longitudinal worker-firm-matched data to examine the relationship between the productivity of workers specifically engaged in innovation and firm size in the pharmaceutical and semiconductor industries. In both industries, we find that inventors' productivity increases with firm size. This result holds across different specifications and even after controlling for inventors' experience, past productivity, the quality of other inventors in the firm, and other firm characteristics.

JEL Classification: O30, O32, O34, J21, J24

Key words: Patents; Innovation; Labor productivity; Research; Firm size

I. Introduction

It has long been recognized that worker wages and possibly productivity is higher in large firms. Moreover, at least since Schumpeter (1942) economists have been interested in the relative efficiency of large firms in the research and development enterprise. This paper examines the relationship between firm size and the productivity of workers specifically engaged in innovation.

This paper exploits panel data on inventors in the pharmaceutical and semiconductor industries, two industries that are prolific generators of innovations and patents. We use patents and patent citations as measures of inventor productivity. We link the inventors to firms in these industries through U.S. patent records, and obtain additional information on both the inventors and their employers from secondary sources. We find that in both industries, inventors' productivity increases with firm size. This result holds across different specifications and even after controlling for inventors' experience, past productivity, the quality of other inventors in the firm, and other firm characteristics.

This paper is organized as follows. The next section summarizes the literatures on the economy-wide productivity-employer size relationship and on the R&D-firm size relationship. Section III describes our empirical method and the dataset we created for this project. Section IV describes our empirical results and Section V concludes.

II. Literature review

Do scientist's productivity vary across firms, and if so, how? The determinants of the productivity of research scientists in firms have received little attention by economists. There are theoretical reasons for and some evidence of systematic variation in the productivity of R&D inputs within firms of differing size, however.

Schumpeter (1942), Panzar and Willig (1981), and Cohen and Klepper (1996), among others, have argued that size may be an important determinant of R&D productivity. Large firms with substantial market share may have an advantage in R&D because monopoly power enables them to capture the returns to innovation. Large firms may also have an advantage because of either scale economies (due to fixed costs to mounting an R&D operation) or because their size affords them wider access to external sources of financing. On the other hand, R&D productivity may in some instances be subject to diseconomies of scale, for example, because of the inflexibility that sets in with the bureaucratization of the R&D enterprise. For any of these reasons, firm size may be a factor in the productivity of research scientists.

There is considerable empirical research examining the relation between the productivity of R&D in firms and firm size. Such research has examined how patent (or citation-weighted patent) yields from R&D activities (usually measured by R&D expenditures) vary with firm size. In examining the relation between size and R&D productivity, researchers face difficulties obtaining accurate measures of R&D expenditures.[1] Many believe that R&D expenditures are especially poorly measured for smaller firms because smaller firms often conduct R&D informally (they lack R&D departments and separate R&D staff). This may explain why many studies fail to show a relationship between size and R&D productivity (see Symeonidis, 1996, for a survey of this literature) or why a few studies even find that the patent yield from R&D expenditures falls with firm size (see, for example, Acs and Audretsch, 1991, Bound et al, 1984, and Hausman et al, 1984). R&D expenditures reflect, of course, not just the human capital

[1] A number of scholars have discussed the strengths and limitations of the various measures of R&D inputs and outputs, including patents for output and R&D expenditures and scientific employment for inputs (e.g., see Cohen and Levin, 1989).

devoted to innovation, but the physical capital devoted to it, as well. To our knowledge no studies in the economic literature have examined the relation between firm size and productivity of scientists.

A substantial literature documents employer size-wage or productivity premium generically, however. Large firms pay workers a premium that is comparable to or even greater than the wage gaps observed between genders, among races, and between unionized and non-unionized workers (see, for example, Oi, 1983; Brown and Medoff, 1989; Davis and Haltiwanger, 1991; and Troske, 1999). Moreover, workers in large firms appear to be more productive. Many hypotheses have been put forward to explain the size-wage/productivity premium. For example, Idson and Oi (1999), Dunne and Schmitz (1992), and others have argued that labor productivity is higher in large firms because large firms adopt new technologies sooner and possess higher quality capital than small firms. In the context of R&D, this would imply that large firms possess newer and more sophisticated laboratories than small firms. Griliches (1970) and Hamermesh (1993) argue that the employer size-productivity premium is due to the complementarity between worker skill and physical capital. Large firms may employ inherently more able workers and impose higher work standards than small firms (see Idson and Oi).

III. Empirical Implementation

3.1 Model Specification

We use panel data on a sample of research scientists in industry to test whether scientists are more productive in large firms. We estimate the effect of firm size on the scientist's labor productivity in patenting using a negative binomial maximum-likelihood regression model

(Hausman, Hall, and Griliches, 1984). We employ a negative binomial model because the number of patents granted to a scientist in a particular year is a nonnegative count variable.[2] We assume that the expected number of patents invented by a scientist, conditional on his/her characteristics, is

$$E(PAT_{it}) = \exp[\alpha + \beta_1 \ln(R\&D_{it}) + \beta_2 \, PHDEG_i + \beta_3 \, EXP_{it} + \beta_4 \, EXP_{it}^2 + \gamma \, X_{it}],$$

where PAT_{it} is the number of patents granted to scientist i that were applied for in year t, and $R\&D_{it}$ is year t R&D expenditures (deflated by the GNP deflator) of the firm to which scientist i's patents are assigned. Note that we use as a measure of firm size the magnitude of research expenditures instead of a more comprehensive firm size measure (such as firm sales) because it is likely a more relevant size factor in a scientist's productivity. To check the robustness of our result, we employ the total sales (SALES) and the total number of employees (EMPLOYEE) as alternative measures for firm size. Note also that we match a scientist with the firm to which the scientist's patents are assigned under the presumption that patents are assigned to inventors' employers. $PHDEG_i$ is a binary variable for those scientists with a Ph.D. degree, EXP_{it} is the number of years elapsed in year t since scientist i had the first patent granted, and EXP_{it}^2 is a squared term. Following the Mincerian earnings regression studies, the two variables, EXP_{it} and EXP_{it}^2, are included to capture the scientist's experience in research. X_{it} is a vector of the characteristics of the firm in year t which scientist i's patents are assigned to. The vector includes the capital-labor ratio (K/L), the inventor-R&D ratio (INV/R&D), the share of Ph.D. degree holders among all patenting inventors in the firm (PHD/INV), the mean experience of the patenting inventors (MEXP), firm age (FIRMAGE), and the number of business lines (NSIC).

---

[2] We prefer the negative binomial specification to a Poisson specification because overdispersion tests indicate that the Poisson assumption that the mean equals the variance is not valid in our data. As shown in Table 1, the mean of our dependent variable, the number of patents, in each industry is smaller than its variance.

Since our basic specification may have individual-specific missing factors that affect the scientist's productivity in patenting, we estimate the negative binomial models with scientist-specific fixed and random effects. We also try a fixed-effects negative binomial model with additional, firm-specific binary variables. Another econometric model applied in our empirical analysis employs ordinary least squares estimation methods, and we try specifications with scientist-specific fixed effects and with between effects. The scientist-specific fixed-effects specification captures "time-series" variability in the regressors, whereas the between-effects specification captures "cross-sectional" variability.

Our data include instances where in a given year a scientist is named as inventor on the patents assigned to multiple firms. These instances can take place, for example, when a scientist changes employers or when he/she is not affiliated with any single firm. In our base specification, we exclude observations of this kind; that is, we exclude observations in which the scientist is matched to multiple firms in one year. In an alternative specification, we include such observations, assuming that the inventor is employed sequentially to each of the firms to which he/she is matched. For these observations, we assume the employment duration with any one firm is proportional to the ratio of patents assigned to that firm to patents assigned to all firms in that year.

3.2 Data and Variables Used

Our data are taken from six sources: (1) Patent Bibliographic data (Patents BIB) released by the U.S. Patent and Trademark Office (USPTO) that contain bibliographic information for U.S. utility patents issued from 1969 to 2002; (2) the Compact D/SEC database which contains firm information taken primarily from 10-K reports filed with the Securities and Exchange Commission; (3) the Standard & Poor's Annual Guide to Stocks-Directory of Obsolete Securities which includes a history of firm ownership changes due to mergers and acquisitions, bankruptcy, dissolution, and name changes; (4) the patent citations data collected by Hall, Jaffe and Trajtenberg (2001) which contain the citation counts per cited patent over the period 1975-1999; (5) the Thomas Register data which report the firm's founding year, and finally (6) the ProQuest Digital Dissertations Abstracts database which contains information on the date, the field, and the type of degree for degree holders We match these data to the inventors in the Patents BIB data by scientists' names.

As the first step for merging these datasets, we choose all firms whose primary SIC is 2834 (pharmaceutical preparation) or 3674 (semiconductor and related devices) in the Compact D/SEC data. We selected these two industries for our study because the firms in these industries are active in patenting and produce homogenous products relative to other industries. Because patents are typically assigned to the firm that employs the inventors, we identify the inventors' employers in the Patents BIB data by patent assignees.

However, because parent firms sometime patent under their own names and other times under the names of their subsidiaries, merging the Patents BIB data with firm-level data in the Compact D/SEC data is not straightforward. Mergers and acquisitions at both the parent firm and subsidiary levels, common in these two industries during the 1990s, and firm name changes further complicate linking the patent to firm-level data. (The USPTO does not maintain a unique identifier for each patenting assignee at the parent firm level nor does it track assignee name changes.) Thus, to use the firm-level information available in the Compact D/SEC data, the names of parent firms and their subsidiaries and the ownership of firms (and/or subsidiaries) must be tracked over the entire period of the study.

To link we use the S&P data to identify whether each assignee in our Patents BIB extract was a stand-alone firm. In some cases, due to a merger or acquisition, for example, the assignee was actually part of another firm at the time of the patent application. In these cases, we substituted this parent firm's name for the name given as the assignee. We then use the subsidiary data in the Compact D/SEC data to track changes in the parent for each of the firms (with the corrected names) in our Patents BIB extract. We use the histories of name changes and M&A's to assign firm identifiers. We thus assign our own firm identifier accordingly after tracking the histories of name changes and M&A's of each firm. For example, if a firm is acquired, we keep separate data on the firm through the year prior to acquisition, because the acquiring firm reports consolidated information and because the patents applied for by the target firm after acquisition should be linked to the acquiring firm. The firm that acquires the target firm retains the same firm identifier. If firms are merged, we keep their observations up to the year prior to merger and assign the newly merged firm a new firm identifier. If a firm changes its name, it retains the same identifier. If a subsidiary's ownership changes, the subsidiary's identifier becomes the identifier of the new parent, from the date of the change forward.

After merging the Patents BIB data with the firm-level information in the Compact D/SEC data, we then link the patent inventors to the firms in the Compact D/SEC data by the final firm name to produce a data set on inventors and patents that includes firm-level data (e.g., R&D expenditures, sales, number of employment) on the patents' assignees.

We obtained the inventors' educational backgrounds (degree types, dates, and fields for those who earned masters or doctoral degrees) from the ProQuest database, and linked this information to the patent and firm data by inventor name. Based on the list of those who earned degrees in all natural science and engineering fields between 1894-2002, the inventors are matched to those on the list by their last, first and middle names.[4]

Information on the number of citations per patent is from Hall, Jaffe and Trajtenberg (2001). These data contain the citation counts for all awarded patents whose application dates precede year 2000. According to Hall et al., 50 percent of all citations to a patent are made at least 10 years after the patent's application date. Thus, for awarded patents applied for after 1993, many citations will occur after 1999. Consistent with this, we observe that the total number of citations for all the patents in our data declines after 1994. We therefore utilize only the data on awarded patents whose application dates fall between the years 1989-1993.

Definitions and summary statistics of variables used in our analysis are reported in Table 1.

IV. Empirical Findings

We estimate the association between the number of an inventor's patent grants applied for in year t (PAT) and his/her employer's R&D expenditures in year t (R&D), our measure of the size of the firm's research enterprise. Tables 2 and 3 report separate estimation results for the pharmaceutical and semiconductor industries, respectively. Our use of contemporaneous R&D, as opposed to lagged R&D, follows the extensive literature estimating patent production

[4] Prior to the matching, we modified persons' names in both datasets by converting all lower case letters to upper case letters, deleted all non-alphanumeric characters, such as commas and hyphens. We noticed that inventors' middle names are sometimes reported inconsistently. That is, sometimes middle names are spelled out, sometimes only their initials are included, and at other times no middle name or initial is included. To achieve more accurate matching of inventors, only the initial was taken as middle name and then inventors with unique last, first and middle name were given a unique identifier. Then the inventors with two or more background information were dropped .

functions (e.g., Hall, Grilliches, and Hausman, 1986). Evidence suggests that R&D activities and innovations occur somewhat simultaneously. Moreover, if a firm attempts to patent an innovation, it files the application while the innovation is being developed or very shortly afterwards (Hall et al.).

Our base model (model 1) includes as regressors two individual-level characteristics in addition to the firm size variable: education and experience.[5] PHDEG is a binary variable for a Ph.D. degree holder. We expect that the majority of the inventors in our data set have at least a college education, and that those with PHDEG=0 have either a bachelor's or a master's degree. Following the Mincerian earnings regression studies, we employ both a linear and a quadratic experience term. In model 2, we include additional firm-level regressors (X) to isolate and differentiate the effect of firm size from those of other firm characteristics.

In models 3 and 4, an alternative measure for firm size is used instead of R&D (SALES in model 3 and EMPLOYEE in model 4, respectively). Model 5 includes the firm's average number of citations per patent (MCITE) as an additional regressor. We did not include this variable among the regressors in models 2-4 because the data on MCITE for the years 1989-93 are assumed not subject to the bias due to citation lag. In model 6, we use the data with monthly observations (see section 3.1 for the data construction).

Models 7 and 8 employ a fixed-effects and a random-effects negative binomial model, respectively, as proposed by Hausman, Hall, and Griliches (1984). Note that "fixed-effects" and "random-effects" in these models apply to the distribution of the dispersion between the conditional mean and the variance of the dependent variable so that the dispersion is the same for the observations in the same group (i.e. the same inventor) but varies from group to group. This

[5] These two factors are frequently shown to be significant in productivity and earnings regressions (e.g. Mincer, 1974, Card, 1999).

method is not equivalent to conventional kinds of fixed-effects or random-effects models in panel data analysis. For instance, the coefficient associated with PHDEG is estimated in model 7 with fixed effects. Coefficients on inventor-specific, time-invariant variables such as PHDEG are not estimable in the conventional fixed-effects models (i.e. models with inventor-specific dummy variables). In model 9, we introduce the conventional inventor-specific fixed-effects in a Poisson model. In model 10, we have both firm-specific fixed-effects and inventor-specific fixed-effects applied to a Poisson model. (Due the computational problem, we haven't been yet able to estimate the negative binomial models with inventor-specific dummy variables. The estimation is in progress.) The last two models use the OLS method with inventor-specific fixed-effects and with between-effects, respectively, to compare the effect of firm size based on within-group variations with that based on between-group variations.

Controlling for the individual characteristics, the results of model 1 in both industries show that there are economies of scale in R&D activity: inventors in larger firms are more productive. The effect of firm size (R&D) on patent productivity (PAT) is in fact significant and positive in all models of both tables. The two other measures for firm size exhibit qualitatively the same effect on PAT. These results support the finding in the labor literature that worker wages and productivity are higher in large firms, but contrasts with findings elsewhere that small firms have higher patent-R&D ratios than large firms. Note that the magnitude of the firm size effect on patenting is generally bigger in the pharmaceutical industry than in the semiconductor industry.[6]

Human capital theory predicts that higher education is associated with higher productivity and that labor productivity has an inverted U-shaped relationship with experience.

[6] Note also that the estimated elasticity of PAT with respect to R&D in model 6 is significantly higher than in other models. This is because the dependent variable in model 6 is the monthly number of patents, whereas it is the annual number of patents in other models.

Both predictions are strongly confirmed in tables 2 and 3. PHDEG exhibits a significant and positive effect on PAT, and EXP has an inverted U-shape relationship with PAT. According to the estimated coefficients in model 1 of tables 1 and 2, the peak in PAT is reached at EXP=16.1 years and 12.3 years in pharmaceutical and semiconductor industry, respectively.

We include the capital labor ratio as a regressor because given R&D expenditures a highly capitalized firm may have a stronger incentive to patent than less capitalized firms. A patent infringement lawsuit that leads to production stoppage will be more destructive for a firm that has made a large capital investment in a state-of-the-art physical plant. Such vulnerability may encourage the firm to develop a diverse portfolio of patents that it can use as a bargaining chip to ward off infringement suits (Cohen et al., 2000; Parr and Sullivan, 1996). In addition, we expect that the capital intensity in the R&D division will be positively correlated with the capital intensity in the entire firm. Therefore, a higher K/L in the entire firm may raise labor productivity in patenting of an inventor due to capital-skill complementarity. The results in both tables show support for this prediction. K/L exhibits a positive effect on PAT when these effects are also statistically significant.

The results in tables 2 and 3 indicate that the patent productivity of an inventor varies positively with the number of patenting inventors in the same firm, holding constant R&D expenditures. This suggests positive productivity spillovers across inventors within a firm.[7] The median experience of inventors in a firm, MEXP, exhibits a positive effect on PAT in general, further evidence that spillover effects exist.

A higher percentage of Ph.D. degree holders among inventors may raise the labor productivity of an inventor in the same firm due to a positive spillover effect. On the other hand,

[7] We did not use INV as a separate regressor along with R&D because of multicollinearity problem. The simple correlation coefficient between these variables is about .9 in our data.

firms with a higher concentration of Ph.D. degree holders may be engaged in innovations with higher economic values and produce a smaller number of patents. The net effect of the proportion of Ph.D. degree holders on labor productivity is thus theoretically ambiguous. Our results show that PHD/INV has a positive, but marginally significant effect on PAT in the pharmaceutical industry, while it has a negative, but marginally significant effect in the semiconductor industry.

Tables 2 and 3 show that the age of a firm is negatively related to the patent productivity of the firm's inventors. Two possible reasons may be that (1) older firms carry out larger scale projects and produce a smaller number of patents with higher economic value, or (2) older firms have exhausted new ideas for innovations and produce fewer patents.

Our results also indicate that the number of business lines in a firm, measured by the number of secondary SIC's classified to the firm (NSIC), has a significantly negative effect on PAT in all models in both tables except the models with fixed effects in the semiconductor industry. Thus, the evidence does not the support the presence of economies of scope in scientific labor productivity. This evidence may reflect instead varying mixes of the technologies researched across firms of varying NSIC. That is, scientists in firms with multiple lines may be working in fields where the economic value of patents is large, relative to the fields in which scientists in firms with few business lines.

V. Conclusion

Our findings can be summarized as follows. Using patents as our measure of a scientist's output, we find that labor productivity of scientists rises with firm size, whether size is measured in R&D expenditures, sales or employees. Our finding that patent productivity rises with firm size even after controlling for measures of the ability of scientists in the firm (including scientist-specific fixed effects), suggest that productivity advantages enjoyed by large firms is not simply due to large firms' ability to hire and retain high quality researchers. In addition, we show that patent productivity of inventors varies positively with the number of patenting inventors in the firm, holding constant R&D expenditures. This suggests positive productivity spillovers across inventors within a firm. Our results are robust across different specifications, including specifications that control for unobserved firm and scientist heterogeneity, and arise in both the pharmaceutical and semiconductor industry analyses.

We make the following two qualifications concerning our analysis, which we hope to remedy in future work. First, patents are an imperfect measure of output because their economic value varies from field to field. We have partly controlled for this by analyzing separately two fairly homogenous industries. Nevertheless, the nature of the innovation and therefore of patents may vary across firms of different sizes within an industry. Thus, in future work, we plan to use citation-weighted patents per inventor as our measure of productivity. Because the number of citations a patent receives is a reflection of the patent's importance, citation-weighted patents are considered a better measure of the economic value of patents by many researchers in the field.

Second, our measure of the patent productivity-firm size relation may be biased if not all inventors in a firm patent each year. Inventors only show up in our data if they invent. Presumably, in any year, firms employ some inventors who do not appear on patents. If small firms in particular employ relatively high numbers of inventors who do not invent then our estimate of the patent productivity-firm size gradient is downward biased. We are presently working on a maximum likelihood estimator that will correct this bias.

References

Acs, Z.J. and Audretsch, D.B. "Innovation, Market Structure, and Firm Size," *The Review of Economics and Statistics*, Vol. 69, No. 4, pp. 567-574, Nov. 1987.

Acs, Z.J. and Audretsch, D.B. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, 78, 678-90, 1988

Baldwin, W.L. and Scott, J.T. *Market Structure and Technological Change.* Chur, Switzerland: Harwood Academic Publishers, 1987

Bound, John, Cummings, Clint, Griliches, Zvi, Hall, Bronwyn H., and Jaffe Adam. "Who Does R & D and Who Patents?" in *R&D, Patents, and Productivity,* ed. Z. Griliches, NBER, 1984.

Brown, Charles and Medoff, James. "The Employer-Size-Wage Effect." *Journal of Political Economy.* Vol. 97, No. 5, pp. 1027-59, October 1989.

Card, David, "The Causal Effect of Education on Earnings," in Handbook of Labor Economics, ed. Orley Ashenfelter and David Card, Elsevier, New York, 1999.

Chandler, A. D. Scale and Scope*: The Dynamics of Industrial Capitalism*, Cambridge, Mass: MIT Press, 1990.

Cohen, W.M. and Klepper, S. "A Reprise of Size and R & D," *The Economic Journal*, Vol. 106, No. 437, pp. 925-951, July 1996.

Cohen, W.M. and Levin, R.C. "Empirical Studies of Innovation and Market Structure," in *Handbook of Industrial Organization,* eds R. Schmalensee and R.D. Willig, New York: North-Holland, 1989

Comanor, W.S. "Research and Technical Change in the Pharmaceutical Industry," *Review of Economics and Statistics,* Vol. 47, pp. 182-90, 1965

Davis, Steven J. and Haltiwanger, John. "Wage Dispersion Between and Within Manufacturing Plants, 1963-1986." In *Brookings Papers on Economic Activity*, pp. 115-200, 1991.

Dunne, Timothy and Schmitz, James A. "Wages, Employment Structure and Employer Size-Wage Premia: Their Relatinship to Advanced-Technology Usage at U.S. Manufacturing Establishments." Discussion paper no. 92-15 Washington, D.C.: Bureau of the Census, Center for Economic Studies (December 1992).

Freeman, C. The Economics of Indsutrial Innovation, 2d. ed. Cambridge, Mass: MIT Press, 1982

Graves, S.B. and Langowitz, N. “Innovative Productivity and Returns to Scale in the Pharmaceutical Industry,” *Strategic Management Journal*, Vol. 14, pp. 593-605, 1993

Griliches, Zvi. “Notes on the Role of Education in production Functions and Growth Accounting.” In W. Lee Hansen (ed.), Education, Income, and Human Capital (New York: NBER, 1970).

Hamermesh, Daniel S. *Labor Demand.* (Princeton, NJ: Princeton University press, 1993.)

Hall, B. H., A. B. Jaffe, and M. Tratjenberg "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498, 2001.

Hausman, Jerry A., Hall, Bronwyn H., and Griliches, Zvi. “Econometric Models for Count Data with an Application to the Patents R and D Relationship,” *Econometrica* 52, 1984.

Idson, Todd L. and Oi, Walter Y. “Workers are More Productive in Large Firms.” *American Economic Review.* Vol 89, No.2, pp. 104-108, May 1999.

Henderson, Rebecca and Cockburn, Ian, “Scale, scope, and spillovers: the determinants of research productivity in drug discovery,” *RAND Journal of Economics*, 27(1), pp. 32-59, Spring 1996.

Jaffe, Adam B., Trajtenberg, Manuel, and Henderson, Rebecca. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *The Quarterly Journal of Economics*, Vol. 108, No. 3, pp. 577-598, August 1993

Jensen, E.J. “Research Expenditures and the Discovery of New Drugs,” *Journal of Industrial Economics*, Vol. 36, pp. 83-95, 1987.

Kleinknecht, A. “Measuring R & D in Small Firms: How Much are we Missing?” *The Journal of Industrial Economics*, Vol. 36, No. 2, pp. 253-256, Dec. 1987.

Mincer, Jacob, *Schooling, Experience, and Earnings,* New York, Columbia University Press, 1974.

Oi, Walter Y. “The Fixed Employment Costs of Specialized labor.” In Jack E. Triplett (ed.), *The Measurement of Labor Costs* (Chicago: University of Chicago Press for NBER, 1983).

Panzar, J.C. and Willig, R.D. “Economies of Scope (in Sustainability Analysis),” *The American Economic Review*, Vol. 71, No. 2, pp. 268-272, May, 1981.

Pavitt, K., Robson, M., and Townsend, J. “The Size Distribution of Innovative Firms in the U.K.: 1945-1983,” *Journal of Industrial Economics*, V. 35, 297-316, 1987.

Scherer, F.M. “R & D and Declining Productivity Growth,” *The American Economic Review*, Vol. 73, No. 2, pp. 215-218, May, 1983

Schwartzman, D. *Innovation in the Pharmaceutical Industry*. Baltimore: Johns Hopkins University Press, 1976

Schumpeter, J. *Capitalism, Socialism, and Democracy*. Harper and Row, New York, 1942.

Troske, Kenneth. "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *The Review of Economics and Statistics.* Vol 81, No. 1, pp. 15-26, February 1999.

Vernon, J.M. and Gusen, P. "Technical Change and Firm Size: The Pharmaceutical Industry," *Review of Economics and Statistics*, Vol. 56, pp. 294-302, 1974

Table 1. Variable Definition and Sample Statistics

	Definition [Data Source]	Mean (Variance) Pharmaceutical	Mean (Variance) Semiconductor
PAT	Number of patents granted per inventor in a firm by application year [Patents BIB]	1.602 (2.339)	1.964 (6.701)
R&D	Real R&D expenditures in 1996 constant dollars [Compact D/SEC]	8,316 (2.151E+7)	5,517 (4.345E+7)
SALES	Real sales volume in 1996 constant dollars [Compact D/SEC]	78,633 (2.504E+9)	53,467 (5.321E+9)
EMPLOYEE	Number of Employees [Compact D/SEC]	35,161 (3.469E+8)	19,275 (3.777E+8)
PHDEG	Binary variable for a Ph.D. degree holder [ProQuest]	0.292 (0.207)	0.125 (0.109)
EXP	Years elapsed since the inventor's first patent granted is applied [Patents BIB]	6.505 (42.691)	4.667 (28.101)
K/L	Capital-labor ratio, or deflated plant and equipment over the number of employees [Compact D/SEC]	913 (0.577)	2.143 (97.354)
INV/R&D	Inventor-R&D ratio, or total number of inventors in a firm by application year over real R&D expenditures [Patents BIB, Compact D/SEC]	34.088 (742.938)	16.717 (612.623)
PHD/INV	Share of inventors who hold Ph.D. degrees [ProQuest, Patents BIB]	0.297 (0.009)	0.134 (0.005)
MEXP	Median experience of all inventors in a firm [Patents BIB]	6.855 (1.363)	4.994 (2.018)
FIRMAGE	Years elapsed since the founding year of a firm [Thomas Register]	92.950 (1387.671)	28.136 (370.651)
NSIC	Number of secondary SIC's assigned to a firm [Compact D/SEC]	4.482 (4.718)	2.320 (2.465)
MCITE	Average number of citations per patent [Citation]	5.470 (6.850)	10.931 (24.514)

Table 2. Scientist's Labor Productivity in Patenting: Pharmaceutical Industry

Dependent Variable: PAT

	Negative Binomial Models					
	(1) Base	(2) More firm controls	(3) Sales as size measure	(4) Employees as size	(5) Citation added	(6) Monthly data
ln(R&D)	0.0103 *2.12*	0.1573 *10.95*			0.1566 *7.52*	0.8281 *8.05*
ln(SALES)			0.0795 *8.46*			
ln(EMPLOYEE)				0.0639 *4.91*		
PHDEG	0.1046 *6.62*	0.0998 *5.61*	0.0992 *5.56*	0.0997 *5.57*	0.0978 *4.26*	0.3502 *3.34*
EXP	0.0450 *13.45*	0.0529 *13.46*	0.0528 *13.38*	0.0525 *13.28*	0.0557 *10.24*	0.2609 *10.72*
EXP^2	-0.0014 *-8.14*	-0.0017 *-8.21*	-0.0017 *-8.14*	-0.0016 *-8.06*	-0.0019 *-6.42*	-0.0085 *-7.81*
ln(K/L)		0.0806 *3.79*	0.1405 *7.01*	0.1691 *8.57*	0.0754 *2.50*	0.4833 *2.99*
ln(INV/R&D)		0.2657 *15.59*	0.1931 *13.93*	0.1534 *11.44*	0.2118 *9.26*	1.4463 *11.09*
ln(PHD/INV)		0.0410 *1.44*	0.0879 *3.17*	0.1168 *4.08*	0.0253 *0.81*	0.2726 *1.05*
ln(MEXP)		0.0817 *1.84*	0.0945 *2.19*	0.1281 *2.96*	0.1515 *3.12*	1.3318 *3.17*
ln(FIRMAGE)		-0.0857 *-6.50*	-0.0576 *-4.53*	-0.0527 *-3.85*	-0.0975 *-5.41*	-0.5720 *-7.14*
ln(NSIC)		-0.1224 *-8.05*	-0.1016 *-6.76*	-0.0755 *-5.14*	-0.1031 *-4.30*	-0.7673 *-6.72*
ln(MCITE)					0.0415 *1.83*	
Observations	30,892	21,212	21,191	21,212	10,341	266,612

Note: t-statistics in *Italic*. Constant terms are not reported.

Table 2. (cont.) Scientist's Labor Productivity in Patenting: Pharmaceutical Industry

Dependent Variable: PAT

	Negative Binomial		Poisson		OLS	
	(7) Fixed Effects	(8) Random Effects	(9) Fixed Effects	(10) Inventor- & Firm specific FE	(11) Fixed Effects	(12) Between Effects
ln(R&D)	0.3581	0.1490	0.4003	0.5197	0.2492	0.0397
	6.70	*11.90*	*7.24*	*7.84*	*7.85*	*6.39*
PHDEG	-13.1319	0.0915				0.0473
	-0.13	*6.85*				*6.55*
EXP	0.0202	0.0487	0.0205	0.0086	0.0153	0.0245
	2.46	*15.93*	*2.45*	*0.93*	*2.95*	*14.24*
EXP^2	-0.0013	-0.0015	-0.0015	-0.0015	-0.0010	-0.0007
	-4.49	*-10.90*	*-5.26*	*-5.26*	*-5.20*	*-9.00*
ln(K/L)	-0.0464	0.0740	-0.0117	0.0164	-0.0287	0.0619
	-0.88	*4.22*	*-0.22*	*0.28*	*-0.84*	*6.76*
ln(INV/R&D)	0.5738	0.2597	0.6350	0.6850	0.3661	0.0662
	12.86	*15.98*	*14.43*	*14.30*	*13.63*	*8.02*
ln(PHD/INV)	-0.0620	0.0363	-0.0387	-0.0191	-0.0557	0.0152
	-0.95	*1.38*	*-0.60*	*-0.28*	*-1.39*	*1.16*
ln(MEXP)	0.5772	0.0836	0.6067	0.6483	0.2614	0.0123
	5.08	*1.84*	*5.38*	*5.57*	*3.82*	*0.55*
ln(FIRMAGE)	-0.3246	-0.0815	-0.2943	-0.2196	-0.1769	-0.0255
	-4.39	*-8.10*	*-3.80*	*-1.75*	*-3.74*	*-4.90*
ln(NSIC)	-0.1829	-0.1267	-0.2049	-0.2136	-0.1021	-0.0253
	-4.99	*-8.96*	*-5.68*	*-5.28*	*-4.33*	*-3.49*
Observations	14,719	21,212	14,719	14,719	21,212	21,212

Note: t-statistics in *Italic*. Constant terms are not reported.

Table 3. Scientist's Labor Productivity in Patenting: Semiconductor Industry

Dependent Variable: PAT

	Negative Binomial Models					
	(1) Base	(2) More firm controls	(3) Sales as size measure	(4) Employees as size	(5) Citation added	(6) Monthly data
ln(R&D)	0.0316	0.1497			0.0834	0.6529
	6.39	*15.13*			*5.77*	*9.84*
ln(SALES)			0.1202			
			13.39			
ln(EMPLOYEE)				0.1211		
				11.21		
PHDEG	0.1239	0.1247	0.1260	0.1229	0.1287	0.2887
	2.50	*2.61*	*2.65*	*2.59*	*1.46*	*2.23*
EXP	0.1086	0.1123	0.1120	0.1121	0.0986	0.3046
	15.43	*16.86*	*16.88*	*16.84*	*8.03*	*15.36*
EXP^2	-0.0044	-0.0044	-0.0044	-0.0044	-0.0049	-0.0112
	-12.12	*-13.20*	*-13.27*	*-13.24*	*-7.04*	*-11.33*
ln(K/L)		0.0312	0.0350	0.0632	0.1632	0.0917
		3.47	*3.68*	*6.49*	*4.29*	*2.20*
ln(INV/R&D)		0.4039	0.3456	0.3276	0.2818	1.5484
		17.89	*16.33*	*15.46*	*7.17*	*14.74*
ln(PHD/INV)		-0.0644	-0.0716	-0.0709	-0.0316	-0.1715
		-2.83	*-3.12*	*-3.11*	*-1.24*	*-1.92*
ln(MEXP)		0.0005	0.0286	0.0489	0.0448	0.3653
		0.01	*0.70*	*1.22*	*0.87*	*1.58*
ln(FIRMAGE)		-0.1714	-0.1634	-0.2053	0.0175	-0.5958
		-8.50	*-7.02*	*-7.51*	*0.65*	*-5.36*
ln(NSIC)		-0.0756	-0.0872	-0.0571	-0.1157	-0.3764
		-3.65	*-4.36*	*-2.83*	*-3.39*	*-4.31*
ln(MCITE)					0.1877	
					3.90	
Observations	20,663	17,933	17,933	17,933	4,021	225,270

Note: t-statistics in *Italic*. Constant terms are not reported.

Table 3. (cont.) Scientist's Labor Productivity in Patenting: Semiconductor Industry

Dependent Variable: PAT

	Negative Binomial		Poisson		OLS	
	(7) Fixed Effects	(8) Random Effects	(9) Fixed Effects	(10) Inventor- & Firm specific FE	(11) Fixed Effects	(12) Between Effects
ln(R&D)	0.1625	0.1312	0.2420	0.3445	0.1679	0.0576
	5.14	*16.44*	*6.44*	*6.82*	*5.63*	*11.90*
PHDEG	-0.3515	0.0800				0.0673
	-2.31	*3.55*				*4.71*
EXP	0.0436	0.0829	0.0621	-0.0069	0.0325	0.0492
	4.04	*20.39*	*5.34*	*-0.46*	*3.60*	*17.42*
EXP^2	-0.0003	-0.0031	-0.0008	-0.0008	0.0000	-0.0019
	-0.57	*-15.73*	*-1.94*	*-1.73*	*-0.05*	*-14.05*
ln(K/L)	-0.0204	0.0181	-0.0233	-0.0159	-0.0131	0.0203
	-2.28	*2.89*	*-2.81*	*-1.90*	*-1.81*	*4.13*
ln(INV/R&D)	0.4261	0.3275	0.5979	0.6415	0.3202	0.1537
	11.86	*23.29*	*16.82*	*17.41*	*11.76*	*17.16*
ln(PHD/INV)	-0.0194	-0.0481	-0.0171	-0.0093	-0.0142	-0.0448
	-0.53	*-2.83*	*-0.49*	*-0.26*	*-0.52*	*-3.94*
ln(MEXP)	0.2972	0.1452	0.2624	0.2118	0.1103	0.0715
	3.49	*3.95*	*3.16*	*2.45*	*1.71*	*3.03*
ln(FIRMAGE)	0.0469	-0.1410	-0.1408	0.6407	-0.1168	-0.0665
	0.55	*-8.03*	*-1.47*	*3.59*	*-1.60*	*-6.47*
ln(NSIC)	0.0912	-0.0547	0.1069	0.0768	0.0431	-0.0394
	3.46	*-4.38*	*4.49*	*3.07*	*2.12*	*-4.69*
Observations	11,479	17,933	11,479	11,479	17,933	17,933

Note: t-statistics in *Italic*. Constant terms are not reported.